DOCUMENT RESUME

ED 220 051 HE 015 410

AUTHOR Dumont, Richard G.; And Others
TITLE A Comparison of Methodological Approaches for Evaluating the Quality of Basic Skills Programs. AIR Forum 1982 Paper.
PUB DATE May 82
NOTE 32p.; Paper presented at the Annual Forum of the Association for Institutional Research (22nd, Denver, CO, May 16-19, 1982).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Accountability; *Basic Skills; College Programs; Comparative Analysis; Compensatory Education; *Developmental Studies Programs; *Educational Assessment; Educational Quality; Evaluation Criteria; Higher Education; Institutional Research; Outcomes of Education; *Program Evaluation; Remedial Mathematics; Remedial Programs; *Research Design; Research Methodology; State Universities
IDENTIFIERS *AIR Forum

ABSTRACT

A multiple-indicator and multiple-design approach for evaluating the quality of basic skills programs in reading, writing, and mathematics that has been implemented within a statewide system of higher education in the southeastern United States was assessed. Several of the evaluation methodologies are compared in order to determine whether results are mutually supportive or contradictory and whether they are suggestive of interchangeability or complementarity as regards indicators and evaluative designs. Data are derived from a project designed to evaluate the effectiveness of developmental studies programs at a regional university and community colleges. Attention is focused upon a program in remedial mathematics at the regional, public university. The multiple-indicator and multiple-design strategy utilized both short-range and long-range post-program measures. The results of applying the following three designs are considered: single-group pretest-posttest comparison, the remediated-exempted comparison, and the cross-program comparison. Findings indicate that the application of the single-group pretest-posttest comparison suggested program effectiveness. While the findings were less clear, program effectiveness was also generally supported by the remediated-exempted comparison. Implementation of the cross-program comparison provided negative data regarding program performance on both short-range and long-range postprogram measures; however, this finding is qualified. It is suggested that the choice of the evaluative design is likely to have a greater effect upon the results of the evaluation than the choice of either the postprogram measure or the grouping statistic. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED220051

20c(2) 35

A COMPARISON OF METHODOLOGICAL APPROACHES

FOR

EVALUATING THE QUALITY OF BASIC SKILLS PROGRAMS

by

Richard G. Dumont, Ph.D.
Professor of Sociology and Chairman,
Department of Sociology and Philosophy
Tennessee Technological University
P.O. Box 5052
Cookeville, Tennessee 38501
Tel. (615) 528-3208
and
Director of the Student Assessment
and
Program Evaluation Project
The State University and Community College System
of Tennessee
1161 Murfreesboro Road
Nashville, Tennessee 37217
Tel. (615) 741-4821

Mr. James T. Jones
Research Assistant,
Department of Educational Psychology
Tennessee Technological University
Cookeville, Tennessee 38501
Tel. (615) 528-3552

Albert J. Bekus, Ph.D.
Professor of English and Director,
Developmental Studies Program
Austin Peay State University
Clarksville, Tennessee 37040
Tel. (615) 648-7612

*Presented at the Twenty-Second Annual Forum of the Association for Institutional Research, May 16-19, 1982, Denver, Colorado.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Association for

~~Institutional Research~~

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

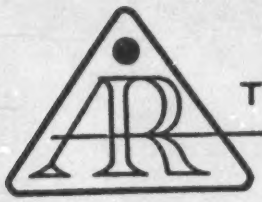
U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

HE 015 410



THE ASSOCIATION FOR INSTITUTIONAL RESEARCH

This paper was presented at the Twenty-Second Annual Forum of the Association for Institutional Research held at the Denver Hilton Hotel in Denver, Colorado, May 16-19, 1982. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum papers.

D. R. Coleman, Chairman
Forum Publication
Advisory Committee

A COMPARISON OF METHODOLOGICAL APPROACHES

FOR

EVALUATING THE QUALITY OF BASIC SKILLS PROGRAMS

(ABSTRACT)

A multiple-indicator and multiple-design approach for evaluating the quality of basic skills programs in reading, writing, and mathematics has been implemented at institutions within a statewide system of higher education in the southeastern United States. Several of the methodological alternatives applied to one of the programs are compared in order to determine whether results are mutually supportive or contradictory and whether they are suggestive of interchangeability or complementarity as regards indicators and evaluative designs.

A COMPARISON OF METHODOLOGICAL APPROACHES FOR EVALUATING THE QUALITY OF BASIC SKILLS PROGRAMS

"Excellence, Equality, and the New Economy: What's Essential?" was the title of the keynote address at the 1982 National Conference on Higher Education of the American Association for Higher Education. The topic could not have been a more timely one. Indeed, in the current inflationary-recessionary period, the tension between the values of quality and equality in American higher education appears to be becoming increasingly problematic. "Project Equality," a nation wide effort sponsored by the College Board, and the Southern Regional Education Board's Task Force Report on Higher Education and the Schools, "The Need for Quality," are but two examples of a growing number of indications that the quality vs. equality dilemma is becoming a major topic of discussion, argumentation and debate among faculties, administrators, governing boards, and state legislatures and executive offices (Casteen, 1982; SREB, 1981).

One can hardly address this sensitive and potentially explosive topic without confronting simultaneously the closely related issue of the need for or place of "developmental studies" or "remedial education" in higher education. For example, the SREB Task Force report considers the challenge of "Finding the most effective setting for remedial studies," and takes the position that, "In the long term, the need for remedial education at the collegiate level should be reduced substantially by measures at elementary and secondary levels to improve student performance" (SREB, 1981, p. 19). The report also includes the formal recommendation that, "The state boards of education and higher education in each state should appoint a joint committee to examine remedial education provided within their state, for the

purpose of determining the most effective settings and models to serve the needs of insufficiently prepared high school graduates" (SREB, 1981, p. 19).

As the various policy alternatives are considered in the several political environments, sound and convincing evidence regarding the effectiveness and efficiency of collegiate remedial instruction programs will be sought. Institutional researchers will be among those whose expertise will be directed toward designing and conducting both formative and summative evaluation studies of remedial programs.

Fortunately, institutional researchers are able to refer to a growing body of literature on both the rationale and the methodology of program evaluation. Representative of recent contributions are the efforts of Anderson and Ball (1978), Anderson, Ball, Murphy, and Associates (1973), Astin (1970a; 1970b; 1974), Baird (1977), Balderston (1974), Bowen (1974; 1977), Cooley (1976), Dressel (1972; 1976), Feldman (1970), Fincher (1978a; 1978b; 1979), Folger (1977), Hartnett (1971), Lenning, Lee, Micek, and Service (1977), Miller (1979), Pedazur (1975), and Rock, Centra, and Linn (1969), to cite but a few. While much of the literature on program evaluation has focused upon the assessment of "general education" and "major field" processes and outcomes, it is also evident that "developmental studies" or "basic skills" programs have become the subjects of serious, scholarly attention (Baldwin and Others, 1975; Carman, 1971; Cosby, 1975; Cross, 1976; Entwistle, 1960; Gordon, 1970; Grant and Hoeber, 1978; Maxwell, 1979; Piesco, 1978; Roueche, 1973, 1977; Sharon, 1970; Trillin and Associates, 1980; and Zwick, 1965).

The purpose of the present study is to identify and to discuss several of the methodological alternatives available for the evaluation of basic skills programs and to illustrate their use by reference to an ongoing evaluation being conducted as part of a statewide system effort in the Southeastern United States. We shall be particularly interested in comparing the

results of several of the methodological alternatives in order to determine whether such results are mutually supportive or contradictory and whether they are suggestive of interchangeability or complementarity as regards one another.

THE METHODOLOGICAL OPTIONS

In their timely and substantial essay, Akst and Hecht (1980) have identified and discussed a variety of methodological options for the evaluation of basic skills programs. When the objective of evaluation is to provide evidence regarding the effectiveness of remedial instruction, the researcher has important decisions to make regarding the selection of preprogram and postprogram measures, the choice of the evaluative design, and the extraction and definition of the grouping statistic. Recognizing the numerous practical, political, and ethical problems associated with the several decision alternatives, Akst and Hecht emphasize the need to be, "objective... informed... comprehensive... pragmatic... politic... selective... and prepared to compromise" (Akst and Hecht, 1980, pp. 294-296).

Selecting the preprogram measure constitutes a very important initial step in the research process. Ordinarily, the preprogram measure involves a test which may be used, "as a base line for later comparisons, a device for sorting out groups to be compared, and a guage of whether groups were equivalent before the start of remedial instruction" (Akst and Hecht, 1980, p. 266). The preprogram measure should satisfy at minimum measurement criteria of validity and reliability as well as those of feasibility and practicality of administration.

The choice of the postprogram measure is especially critical to the outcome of the evaluation. Of particular importance is the fact that

previous research indicates that short-range postprogram measures are likely to be indicative of program success, while long-range measures tend to be more equivocal (Akst and Hecht, 1980, p. 293). Short-range measures may involve end-of-course performance on a test, course grades; or they may involve the use of gain scores: The average score on the pretest is subtracted from the average score on the posttest as a measure of learning growth, when the posttest is either the same or a parallel form of the pretest.

While short-range measures concern assessment of student performance at the end of the course, long-range measures involve instructional effectiveness indicators of student behavior after the remedial program has been completed. Examples of long-range postprogram measures are grade in next related course, grade point average after a specified number of semesters or quarters, accumulated credits, and attrition rates. Postprogram measures should also be selected with a view toward satisfying validity and reliability criteria, and especially with a clear awareness of the fact that the results of any particular evaluation effort may be affected greatly by the type of postprogram measure adopted.

The outcome of evaluation may also be affected by the nature of the evaluative design. Akst and Hecht (1980, pp. 268-294) have described and discussed the respective advantages and disadvantages of eight different designs: the Single-Group Pretest-Posttest Comparison, the Remediated-Unremediated Comparison, the Marginally-Remedial, Marginally-Exempted Comparison, the Regression-Discontinuity Design, the Remediated-Exempted Comparison, the Cross-Program Comparison, the Norm-Group Design, and the Historical Comparison. Each design is characterized in terms of its typical preprogram measure(s), postprogram measure(s), feasibility problems, possible biases, comprehensiveness, equitability, and educational standard. Feasibility problems involve such matters as the ethics of withholding remediation

from students constituting a control group and assembling the control group for a posttest. The possible biases result from tests and testing procedures, but mostly they derive from the type of evaluation design selected. The single-group pretest-posttest comparison involves the largest number of possible biases. These inlcude, "Test administration, student attitudes, teaching to the test, practice effect, instrument, Hawthorne effect, dropout, regression toward the mean, external learning history, and maturation." At the other extreme, involving the fewest number of possible biases, are designs such as cross-program comparison, which is subject to the effects of, "Initial differences, history, and the differential operation of other biases" (Akst and Hecht, 1980, pp. 290-291).

All but one of the eight designs are comprehensive, and only one or two are inequitable; i.e., "results can be interpreted if they go in one direction, but not if they go in the other." Finally, half of the designs described provide an educational standard, some reasonable standard of performance.[1]

The selection of the grouping statistic may also have a determining effect upon the outcome of the evaluation. As Akst and Hecht caution, "It is important that the grouping statistic be chosen at the start of the evaluative study and that it be related to program objectives. Otherwise there may be a great temptation to compute the results in various ways, reporting only the most favorable" (Akst and Hecht, 1980, p. 294). In deciding upon the grouping statistic, the researcher ordinarily selects either an average or a percentage. For example, a short-range measure of program performance might be the average gain score on a test or it might be the percentage of students passing the test. Similarly, a long-range indicator could be the weighted average of the cumulative grade point averages after a specified number of semesters, or it could be the percentage of students with grade point averages of 2.0 or higher.

If there is one message that comes through clearly from a perusal of Akst and Hecht's careful, thorough and systematic methodological survey, it is that there is no single measure, design, or grouping statistic which may be employed to provide conclusive evidence regarding program effectiveness or ineffectiveness. Appreciation of the several methodological options and problems has clear implications for an informed approach to the evaluation of basic skills programs. With regard to the selection of program measures and grouping statistics, a multiple-indicator strategy is implied. Several, rather than only one, indicators of program effectiveness should be considered. Similarly, since each evaluative design is subject to its own peculiar configuration of weaknesses, a multiple-design strategy is also a desideratum. As Akst and Hecht have stated, "Combining designs can only give a more rounded picture of program effectiveness" (Akst and Hecht, 1980, p. 294).

DATA

Our data derive from a project designed to evaluate the effectiveness of developmental studies programs in composition, mathematics and reading within a state-wide system of higher education in the Southeastern United States. The project is sponsored by the system governing board, and it was initiated in the fall of 1979 as a pilot effort involving one of the system's regional universities and one of its community colleges. The project was expanded in the fall of 1981 to include two additional community colleges.

At each participating institution, a multiple-indicator and multiple-design strategy has been adopted, utilizing both short- and long-range post-program measures. Accordingly, data are available which allow exploration of the methodological questions posed in the present study. In order to

illustrate aspects of the application of a comprehensive evaluation strategy to a single program, and in the interest of a reasonably parsimonious presentation, we will focus upon the remedial mathematics program at the regional university. In particular, we shall present and discuss the results of the implementation of three designs, designs which differ significantly from one another in terms of such important considerations as postprogram measures, feasibility problems, and possible biases. Those designs are the Single-Group Pretest-Posttest Comparison, the Remediated-Exempted Comparison, and the Cross-Program Comparison, and their distinguishing characteristics are highlighted in Table 1.

Although it is clearly the simplest and the most feasible, there is little doubt that the Single-Group Pretest-Posttest Comparison is the least adequate of the three designs. Its inadequacy is due to the fact that it is ordinarily restricted to a short-range postprogram measure (test) and also because of the absence of a control group which would allow the elimination or estimation of numerous possible biases. The Cross-Program Comparison is one of the "strongest" designs available to the researcher, since many biases are potentially eliminated because of the availability of an appropriate comparison group. Another strength of this design is that the utilization of both short-range and long-range postprogram measures is possible. The design is rarely feasible, however, due to the difficulties of finding comparable and cooperating institutions. Fortunately, our data allow us to examine the application of this design, since it was implemented simultaneously at one of the system's community colleges. Remediated students at the community college were defined, selected, and tested utilizing the same procedures and measures as those used at the regional university. In fact, the faculties of the two institutions developed the pretest and posttest instruments as a joint,

collaborative effort, thus allowing this rather unique opportunity to implement the inter-institutional comparison.

The Remediated-Exempted Comparison (like the Marginally-Remedial, Marginally-Exempted Comparison and several other designs) is an intermediate or compromise alternative as compared with the two discussed above. Unlike the Single-Group Pretest-Posttest Comparison, it is not restricted to short-range postprogram measures, and its use of a comparison group allows for the elimination of several possible biases. Compared with the Cross-Program Comparison, however, the Remediated-Exempted design is not equitable and the educational standard may be too stringent.

We shall examine the results of the application of the three designs, proceeding from the weakest to the strongest, and paying particular attention to what additional information is provided concerning program effectiveness and the the matter of consistency or inconsistency of findings.

FINDINGS

Selected statistics describing the Single-Group Pretest-Posttest Comparison are presented in Table 2. The data pertain to 67 first-time freshmen enrolling at the regional university during the fall quarter of 1979. The students were the subjects of remediation because their score on the ACT Math Subtest was less than 14 and/or because they scored less than 70 on a faculty-developed math test. Preprogram measures are quite clearly descriptive of students deficient in college-level mathematical ability. Their average score on the ACT Math Subtest was 11.25, well below national, regional, and institutional means. Similarly, the average score on the faculty-developed math test was 45.83, significantly below the passing score of 70.

Readministration of the faculty-developed math test at the end of the fall quarter remedial course, together with course grades, provide the bases for the evaluation of program effectiveness. Comparison of posttest to pretest scores would seem highly suggestive of greatly improved mathematical proficiency on the part of the remediated students. The posttest average score of 65.67 represents a marked and statistically significant gain over the pretest. And, utilizing Bowen's suggested, empirically-based typology for the characterization of change (1977, p. 103), we would describe the gain (as expressed in standard deviation units) as reflecting "a very large increase" in performance. The average of the grades received in the remedial math course are consistent with posttest results in supporting claims of program effectiveness, with a mean grade of 2.32 indicative of a 'C' average.

Shifting our focus from averages to percentages as grouping statistics provides similar and consistent indications of program success. The pretest to posttest increase in the percent of students passing the faculty-developed math test was 40% and the percent of students receiving grades of 'C' and higher was 55%.

If evaluation efforts were to cease with the implementation of the Single-Group Pretest-Posttest Comparison, the available evidence would indeed be most suggestive of an effective remedial program. Recognizing the limitations of this evaluative design, however, one is prone to raise important questions, such as, "What happens to students after they complete the remedial course? How do they perform in non-remedial college math courses? And, how do they compare with non-remedial students in terms of course and overall performance? That is, does remediation help to bridge the preparedness gap between those who were and those who were not deficient in mathematical ability upon first enrollment at the university?"

The Remediated-Exempted Comparison allows at least partial answers to the above questions. Table 3 displays selected statistics describing the Remediated-Exempted Comparison. The 67 remediated students are compared with those 55 first-time freshmen who enrolled in the first required college algebra course during the fall quarter of 1979. As their ACT Math Subtest score reveals, the students were significantly better prepared in their ability to handle college-level math, with an average score of 20.29 as compared with 11.25 for the remedial group.

Since the exempted students at the regional university were not required to take the faculty-developed math test, comparison with the exempted students is restricted to long-range postprogram measures. Long-range postprogram measures are of two major types, enrollment and performance in subsequent math courses, and grade point averages and retention rates. The data which pertain to students' experiences with subsequent math courses provide the first negative, contradictory indications concerning program effectiveness. Whereas we have previously observed that 55% of remediated students received satisfactory grades in the remedial course, only 27% enrolled in the first of two required college algebra courses in the subsequent quarter. Similarly, while 56% of the exempted students went on to enroll in the second required college algebra course, the corresponding percentage for the remediated students was 18%.

While the data pertaining to the fact of enrollment in subsequent courses tend to be negative, performance statistics are generally more positive. Although the average grades received by the remediated students in subsequent math courses are slightly less than those of the exempteds, 'C' averages are maintained and the between-group differences are not statistically significant. The near-equal performance of the remediateds as compared with the exempteds

is also supported when the grouping statistic is the percent of satisfactory grades received.

Examination of data regarding the long-range postprogram measures of grade point averages and retention rates also points toward generally positive results. While comparison of the grade point averages for the fall quarter of 1979 through the fall quarter of 1980 reveals superior and statistically significant performance on the part of the exempted students, the remediated students did maintain 'C' averages. Comparison of the percent of satisfactory grade point averages and the retention rates similarly reveal only minor differences between the two groups.

Having examined and considered the evidence pertaining to the Remediated-Exempted Comparison, one is left with somewhat conflicting indications about the issue of program effectiveness. While remediated students compare quite favorably to the exempted students in terms of grade point averages, retention rates and grades in subsequent math courses, the discovery that the remediateds tend to delay or avoid enrollment in the required college-level math courses must be viewed as problematic. Had the remediateds compared favorably to the exempteds in terms of the percentages enrolling in the subsequent courses all of our data would have been consistent and collectively supportive of program effectiveness. This particular finding leads to equivocation, however, and it illustrates dramatically a major weakness of the Remediated-Exempted Comparison. That is, it is inequitable.

The Cross-Program Comparison is an evaluative design which overcomes several of the major problems of both the Single-Group Pretest-Posttest Comparison and the Remediated-Exempted Comparison. It allows an answer to the question, "How does the remedial mathematics program at the university compare with other similar programs at other institutions?" Table 4 displays

selected statistics describing the Cross-Program Comparison between the regional university and the community college. As discussed above, common procedures were used to define, select, and test remediated students at the community college. The fact that the two groups are comparable in terms of mathematical ability prior to remediation is strongly suggested by scores on the preprogram measures. Both the ACT Math Subtest Scores and the pretest scores on the faculty-developed math test are nearly identical for the two groups. For the 67 remediated university students the average score on the ACT Math Subtest was 11.25, and for the 53 remediated community college students it was 11.28. Even the standard deviations are almost identical, differing by only .07. Similarly, the averages for both groups on the faculty-developed math test are nearly identical, 45.83 and 43.28 for the university and community college students, respectively.

Significant differentials in program effectiveness are strongly suggested by a comparison of the two remediated groups on all of the short-range postprogram measures. The average score on the posttest was 65.67 for the university students and 71.72 for the community college students, representing average gains of 19.84 and 28.44, respectively. These findings are also in evidence when the grouping statistics are the percent of satisfactory scores on the posttest and the increase in the percent of satisfactory scores. The superior performance of the community college students is also suggested by a comparison of the average grades and the percent of satisfactory grades in the remedial courses. For example, while 55% of the remediated university students received grades of 'C' or higher in the remedial course, the corresponding percentage for the remediated community college students was 85%.

Comparison of the remedial programs on the basis of the several long-range postprogram measures is also rather consistently suggestive of greater

program effectiveness at the community college. For example, 47% of the remediated students at the community college enroll in their required college-level math course in the subsequent quarter, as compared with the 27% for the university students. While those university students who do enroll in the required math course (a more "selected few") receive grades which are slightly higher on the average than those for the community college students, 2.00 and 1.79, respectively, the differences are not statistically significant. Neither are the differences in the percent of satisfactory grades, 72% and 60%, which point in the same direction as the grade averages.

The long-range postprogram measures of grade point averages and retention rates are similarly consistent with one another and with the short-range postprogram measures in suggesting a more effective remedial mathematics program at the community college.

DISCUSSION

The purpose of the present study was to compare several methodological approaches for the evaluation of basic skills programs. Our focus was upon a program in remedial mathematics at a regional, public university in the Southeastern United States. The multiple-indicator and multiple-design evaluation strategy implemented there and at a community college with a similar remedial program allowed us to compare the results of three evaluative designs together with several short-range and long-range postprogram measures. Our analysis revealed that the application of the Single-Group Pretest-Posttest Comparison suggested program effectiveness. While the findings were somewhat more equivocal, program effectiveness was also generally supported by the Remediated-Exempted Comparison. Implementation of

the Cross-Program Comparison provided negative data regarding program performance on both short-range and long-range postprogram measures, however. Our research also revealed a notable consistency as between the utilization of the short-range and long-range measures of postprogram effectiveness across the several designs. A similar consistency was observed for the various grouping statistics.

Our experience suggests that the choice of the evaluative design is likely to have a greater effect upon the results of the evaluation than the choice of either the postprogram measure or the grouping statistic.

NOTES

1. A detailed and thorough description of the characteristics, strengths and weaknesses of each of these designs is beyond the scope of this study. The reader is referred to the original source (Akst and Hecht, 1980, pp. 290-291).

Table 1: Distinguishing Characteristics of the Single-Group Pretest-Posttest Comparison, the Remediated-Exempted Comparison, and the Cross-Program Comparison.

Design	Preprogram Measure	Postprogram Measure	Feasibility Problems	Possible Biases	Comprehensive?	Equitable?	Provides an Educational Standard?
Single-Group Pretest-Posttest Comparison	Test	Must be comparable to preprogram measure		Test administration, student attitude, teaching to the test, practice effect, instrument, Hawthorne effect, dropout, regression toward the mean, external learning, history, maturation	yes	yes	no
Remediated-Exempted Comparison	Test (or other placement criteria)	GPA, grade in next related course, retention or graduation rate	Assembling exempted students for posttest	Test administration, teaching to the test, student attitude, Hawthorne effect, differential operation of other biases	yes	no	yes (too stringent in some cases)
Cross-Program Comparison	Test	Test GPA, grade in next related course, retention or graduation rate	Finding comparable cooperating colleges	Initial differences, history, differential operation of other biases	yes	yes	yes

Source: Akst, G. and Hecht, M. Program Evaluation. In A.S. Trillin and Associates, Teaching Basic Skills in College. San Francisco: Jossey-Bass, 1980, pp. 290-291.

Table 2: Selected Statistics Describing the Single-Group Pretest-Posttest Comparison at the Regional University for Students Enrolling as First-Time Freshmen During the Fall Quarter, 1979.

Selected Statistics	Values Pertaining to 67 Remediated Students
PREPROGRAM MEASURES	
ACT Math Subtest Scores	
Maximum Score	20
Minimum Score	2
Average Score	11.25
Standard Deviation	4.29
Pretest Scores on Faculty-Developed Math Test	
Maximum Score	74
Minimum Score	14
Average Score	45.83
Standard Deviation	14.39
Percent of Satisfactory Scores (70 and higher)	4%
SHORT-RANGE POSTPROGRAM MEASURES	
Posttest Scores on Faculty-Developed Math Test	
Maximum Score	94
Minimum Score	16
Average Score	65.67
Standard Deviation	17.01
Percent of Satisfactory Scores (70 and higher)	44%
Average Gain on Faculty-Developed Math Test	
Actual	19.84*
In Standard Deviation Units	1.38
Increase in Percent of Satisfactory Scores	40%

Table 2 (continued): Selected Statistics Describing the Single-Group Pretest-Posttest Comparison at the Regional University for Students Enrolling as First-Time Freshmen During the Fall Quarter, 1979.

Selected Statistics	Values Pertaining to 67 Remediated Students
Grades Received in the Remedial Math Course	
**Average Grade (A=4, B=3, C=2, D=1, F=0)	2.32 (1.52)
Percent of Satisfactory Grades (A, B, and C)	55%

* $p < .0001$

** Standard Deviation in Parentheses.

Table 3: Selected Statistics Describing the Remediated-Exempted Comparison at the Regional University for Students Enrolling as First-Time Freshmen During the Fall Quarter, 1979.

Selected Statistics	Values Pertaining to: 67 Remediated Students	55 Exempted Students	Differences
PREPROGRAM MEASURE:			
ACT Math Subtest Scores			
Maximum Score	20	32	-12
Minimum Score	2	14	-12
Average Score	11.25	20.29	- 9.04 (p=.0001)
Standard Deviation	4.29	4.83	- 0.59
LONG-RANGE POSTPROGRAM MEASURES: ENROLLMENT AND PERFORMANCE IN SUBSEQUENT RELATED COURSES			
Percent of Students Enrolling in First Required (Non-Remedial) College Algebra Course	27%	100%	-73% (p=.001)
Performance in First Required College Albegra Course			
*Average Grade (A=4, B=3, C=2, D=1, F=0)	2.00 (1.28)	2.30 (1.37)	- 0.30
Percent of Satisfactory Grades (A, B, and C)	72%	65%	+ 7%
Percent of Students Enrolling in Second Required (Non-Remedial) College Algebra Course	18%	56%	-38% (p=.001)
Performance in Second Required College Algebra Course			
*Average Grade (A=4, B=3, C=2, D=1, F=0)	2.42 (1.08)	2.86 (1.12)	- 0.44
Percent of Satisfactory Grades (A, B, and C)	75%	84%	- 9%

Table 3 (continued): Selected Statistics Describing the Remediated-Exempted Comparison at the Regional University for Students Enrolling as First-Time Freshmen During the Fall Quarter, 1979.

Selected Statistics	Values Pertaining to: 67 Remediated Students	55 Exempted Students	Differences
LONG-RANGE POSTPROGRAM MEASURES: CUMULATIVE GRADE POINT AVERAGES AND RETENTION RATES			
Cumulative Grade Point Averages Fall, 1979 through Fall, 1980			
*Weighted Averages of Cumulative Grade Point Averages:			
Fall, 1979	2.27 (1.11)	2.65 (.96)	- 0.38 (p=.025)
Winter, 1980	2.22 (.95)	2.62 (.92)	- 0.40 (p=.025)
Spring, 1980	2.37 (.83)	2.78 (.82)	- 0.41 (p=.01)
Fall, 1980	2.38 (.80)	2.72 (.93)	- 0.34 (p=.05)
Percent of Satisfactory Cumulative Grade Point Averages (2.0 and higher):			
Fall, 1979	77%	87%	-10%
Winter, 1980	78%	88%	-10%
Spring, 1980	87%	91%	- 4%
Fall, 1980	87%	89%	- 2%
Retention Rates Fall, 1979 through Fall, 1980			
Percentages of Fall Quarter, 1979 First-Time Freshmen Completing:			
Fall Quarter, 1979	97%	96%	+ 1%
Winter Quarter, 1980	88%	94%	- 6%
Spring Quarter, 1980	79%	86%	- 7%
Fall Quarter, 1980	69%	80%	-11%

*Standard Deviation in Parentheses.

Table 4: Selected Statistics Describing the Cross-Program Comparison Between the Regional University and the Community College for Students Enrolling as First-Time Freshmen During the Fall Quarter, 1979.

Selected Statistics	Values Pertaining to: 67 Remediated University Students	53 Remediated Community College Students	Differences	
PREPROGRAM MEASURES:				
ACT Math Subtest Scores				
Maximum Score	20	19	1	
Minimum Score	2	2	0	
Average Score	11.25	11.28	- 0.03	
Standard Deviation	4.29	4.22	0.07	
Pretest Scores on Faculty-Developed Math Test				
Maximum Score	74	66	8	
Minimum Score	14	24	-10	
Average Score	45.83	43.28	2.55	
Standard Deviation	14.39	10.77	3.62	
Percent of Satisfactory Scores (70 and higher)	4%	0%	4%	
SHORT-RANGE POSTPROGRAM MEASURES:				
Posttest Scores on Faculty-Developed Math Test				
Maximum Score	94	96	- 2	
Minimum Score	16	28	-12	
Average Score	65.67	71.72	- 6.05	(p=.05)
Standard Deviation	17.01	15.68	1.33	
Percent of Satisfactory Scores (70 and higher)	44%	67%	-23%	(p=.05)
Average Gain on Faculty-Developed Math Test				
Actual	19.84	28.44	- 8.60	(p=.005)
In Standard Deviation Units	1.38	2.64	- 1.26	
Increase in Percent of Satisfactory Scores	40%	67%	-27%	(p=.02)

Table 4 (continued): Selected Statistics Describing the Cross-Program Comparison Between the Regional University and the Community College for Students Enrolling as First-Time Freshmen During the Fall Quarter, 1979.

Selected Statistics	Values Pertaining to: 67 Remediated University Students	53 Remediated Community College Students	Differences	
Grades Received in the Remedial Math Course				
*Average Grade (A=4, B=3, C=2, D=1, F=0)	2.32 (1.52)	2.85 (1.33)	- 0.53	(p=.05)
Percent of Satisfactory Grades	55%	85%	-30%	(p=.001)
LONG-RANGE POSTPROGRAM MEASURES: ENROLLMENT AND PERFORMANCE IN NEXT RELATED COURSE				
Percent of Students Enrolling in Next Required (Non-Remedial) Math Course	27%	47%	-20%	(p=.05)
Performance in Next Required (Non-Remedial Math Course				
*Average Grade (A=4, B=3, C=2, D=1, F=0)	2.00 (1.28)	1.79 (1.29)	0.21	
Percent of Satisfactory Grades (A, B, and C)	72%	60%	12%	
LONG-RANGE POSTPROGRAM MEASURES: CUMULATIVE GRADE POINT AVERAGES AND RETENTION RATES				
Cumulative Grade Point Averages Fall, 1979 through Fall, 1980				
*Weighted Averages of Cumulative Grade Point Averages:				
Fall, 1979	2.27 (1.11)	2.67 (.62)	- 0.40	(p=.01)
Winter, 1980	2.22 (.95)	2.52 (.69)	- 0.30	(p=.05)
Spring, 1980	2.37 (.83)	2.57 (.62)	- 0.20	
Fall, 1980	2.38 (.80)	2.56 (.62)	- 0.18	

Table 4 (continued): Selected Statistics Describing the Cross-Program Comparison Between the Regional University and the Community College for Students Enrolling as First-Time Freshmen During the Fall Quarter, 1979.

Selected Statistics	Values Pertaining to: 67 Remediated University Students	53 Remediated Community College Students	Differences	
Percent of Satisfactory Cumulative Grade Point Averages (2.0 and higher):				
Fall, 1979	77%	96%	-19%	(p=.01)
Winter, 1980	78%	90%	-12%	(p=.05)
Spring, 1980	87%	96%	- 9%	
Fall, 1980	87%	97%	-10%	
Retention Rates Fall, 1979 through Fall, 1980				
Percentages of Fall Quarter, 1979 First-Time Freshmen Completing:				
Fall Quarter, 1979	97%	100%	- 3%	
Winter Quarter, 1980	88%	94%	- 6%	
Spring Quarter, 1980	79%	89%	-10%	
Fall Quarter, 1980	69%	72%	- 3%	

*Standard Deviation in Parentheses.

REFERENCES

Akst, G. and Hecht, M. Program Evaluation. In A.S. Trillin and Associates, Teaching Basic Skills in College. San Francisco: Jossey-Bass, 1980.

Anderson, S. and Ball, S. The Profession and Practice of Program Evaluation. San Francisco: Jossey-Bass, 1978.

Anderson, S., Ball, S., Murphy, R.T. and Associates. Encyclopedia of Educational Evaluation. San Francisco: Jossey-Bass, 1973.

Astin, A.W. The Methodology of Research on College Impact, Part One. Sociology of Education. 1970a, 43, 223-254.

Astin, A.W. The Methodology of Research on College Impact, Part Two. Sociology of Education. 1970b, 43, 437-450.

Astin, A.W. Measuring the Outcomes of Higher Education. In H.R. Bowen (ed.). Evaluating Institutions for Accountability: New Directions for Institutional Research (number 1). San Francisco: Jossey-Bass, 1974, pp. 23-46.

Baird, L.L. Assessing Student Academic and Social Progress: New Directions for Community Colleges (number 18). San Francisco: Jossey-Bass, 1977.

Balderston, F.E. Managing Today's University. San Francisco: Jossey-Bass, 1974.

Baldwin, J., and others. Survey of Developmental Mathematics Courses at Colleges in the United States. Educational Resources Information Center, ED 125 688. Garden City, N.Y.: American Mathematical Association of Two-Year Colleges, 1975.

Bowen, H.R. (ed.). Evaluating Institutions for Accountability: New Directions for Institutional Research (number 1). San Francisco: Jossey-Bass, 1974.

Bowen, H.R. Investment in Learning. San Francisco: Jossey-Bass, 1977.

Carman, R. A Cost-Effectiveness Analysis of Various Methods of Instruction

in Developmental Mathematics. Educational Resources Information Center, ED 057 793. Santa Barbara, Calif.: Santa Barbara City College, 1971.

Casteen, J.T. Reforming the School-College Curriculum. Atlanta, Georgia: Southern Regional Education Board, Regional Spotlight Vol. XIII, No. 4, January, 1982.

Cooley, W.W., and Lohnes, P.R. Evaluation Research in Education. New York: Irvington Publishers, 1976.

Cosby, J.P. Remedial Education - Is It Worth It? Educational Resources Information Center, ED 099 067. Fort Lauderdale, Fla.: Nova University, 1975.

Cross, K.P. Accent on Learning: Improving Instruction and Reshaping the Curriculum. San Francisco: Jossey-Bass, 1976.

Dressel, P.L. Handbook of Academic Evaluation. San Francisco: Jossey-Bass, 1976.

Dressel, P.L. Institutional Research in the University: A Handbook. San Francisco: Jossey-Bass, 1972.

Entwistle, D. Evaluations of Study Skills Courses: A Review. Journal of Educational Research, 1960, 53, 117-125.

Feldman, K.A. Research Strategies in Studying College Impact. Iowa City, Iowa: The American College Testing Program, Research Report No. 34, May, 1970.

Fincher, C. Economic and Sociological Studies of Educational Effects. The Educational Forum, 1979, January, 139-151.

Fincher, C. Importance of Criteria for Institutional Goals. New Directions for Institutional Research (number 19). San Francisco: Jossey-Bass, 1978b.

Fincher, C. Program Monitoring in Higher Education. New Directions for Program Evaluation (number 3). San Francisco: Jossey-Bass, 1978a.

Folger, J.K. (ed.). Increasing the Public Accountability of Education. New Directions for Institutional Research (number 16). San Francisco: Jossey-Bass, 1977.

Gordon, E.W. Compensatory Education: Evaluation in Perspective. IRCD Bulletin. New York: Columbia University Press, 1970.

Grant, M.K. and Hoeber, D.R. Basic Skills Programs: Are They Working? AAHE-ERIC Higher Education Research Report No. 1. Washington, D.C.: American Association for Higher Education, 1978.

Hartnett, R.T. Accountability in Higher Education: A Consideration of Some of the Problems of Assessing College Impacts. New York: College Entrance Examination Board, 1971.

Lenning, O.T., Lee, Y.S., Micek, S.S., and Service, A.L. A Structure for the Outcomes of Postsecondary Education. Boulder, Colorado: National Center for Higher Education Management Systems, 1977.

Maxwell, M. Improving Student Learning Skills: A Comprehensive Guide to Successful Practices and Programs for Increasing the Performance of Under-prepared Students. San Francisco: Jossey-Bass, 1979.

Miller, R.I. The Assessment of College Performance. San Francisco: Jossey-Bass, 1979.

Pedazur, E.J. Analytic Methods in Studies of Educational Effects. In F.N. Kerlinger (ed.), Review of Research in Education, Volume 3. Itasco, Illinois: F.E. Peacock Publishers, 1975.

Piesco, J. Research on Remedial Programs at the City University of New York: A Review of Evaluative Studies. Unpublished Report, Office of Institutional Research and Analysis, City University of New York, 1978.

Rock, D.A., Centra, J.A., and Linn, R.L. The Identification and Evaluation of College Effects on Student Achievement. Princeton, New Jersey: Educational Testing Service, 1969.

Roueche, J.E. Accommodating Individual Differences. Community College Review, 1973, 1, 24-29.

Roueche, J.E. (ed.). Increasing Basic Skills by Developmental Studies. New Directions for Higher Education. San Francisco: Jossey-Bass, 1977.

Sharon, A.T. Effectiveness of Remediation in Junior College. College Entrance Examination Board Research and Development Report RDR-70-71, No. 2. Princeton, N.J.: Educational Testing Service, 1970.

Southern Regional Education Board. The Need for Quality. Atlanta, Georgia: Southern Regional Education Board, A Report by The Task Force on Higher Education and the Schools, June, 1981.

Trillin, A.S. and Associates. Teaching Basic Skills in College. San Francisco: Jossey-Bass, 1980.

Zwick, E.J. An Evaluation of the Effectiveness of the Remedial Mathematics Program at the Ohio State University. Unpublished doctoral dissertation, Ohio State University, 1964. Dissertation Abstracts, 1965, 25, 6323.